

Dépasser les limites du SNDS via l'intelligence artificielle : une étude de cas via l'étude EpiCART sur 3 hémopathies malignes



Benjamin Grenier¹, Marc Maynadié², Isabelle Borget³, Margaux Berthet⁴, Morgan Dupuis⁴, Pierre Lauvray⁴, Hélène Denis¹, Guillaume Cartron⁵, Nicolas Boissel⁶

¹Heva, Lyon, ²Faculté des Sciences de Santé, Dijon, France, ³Études et Recherche en Économie de la Santé, Institut Gustave Roussy, Paris, ⁴Gilead Sciences, Boulogne Billancourt, ⁵Centre Hospitalier Lapeyronie, CHU Montpellier, Montpellier, ⁶Hôpital Saint Louis, AP-HP, Paris

Colloque Données de Santé en Vie Réelle (DSVR) 2023 | 20 juin 2023, Paris, France

467388

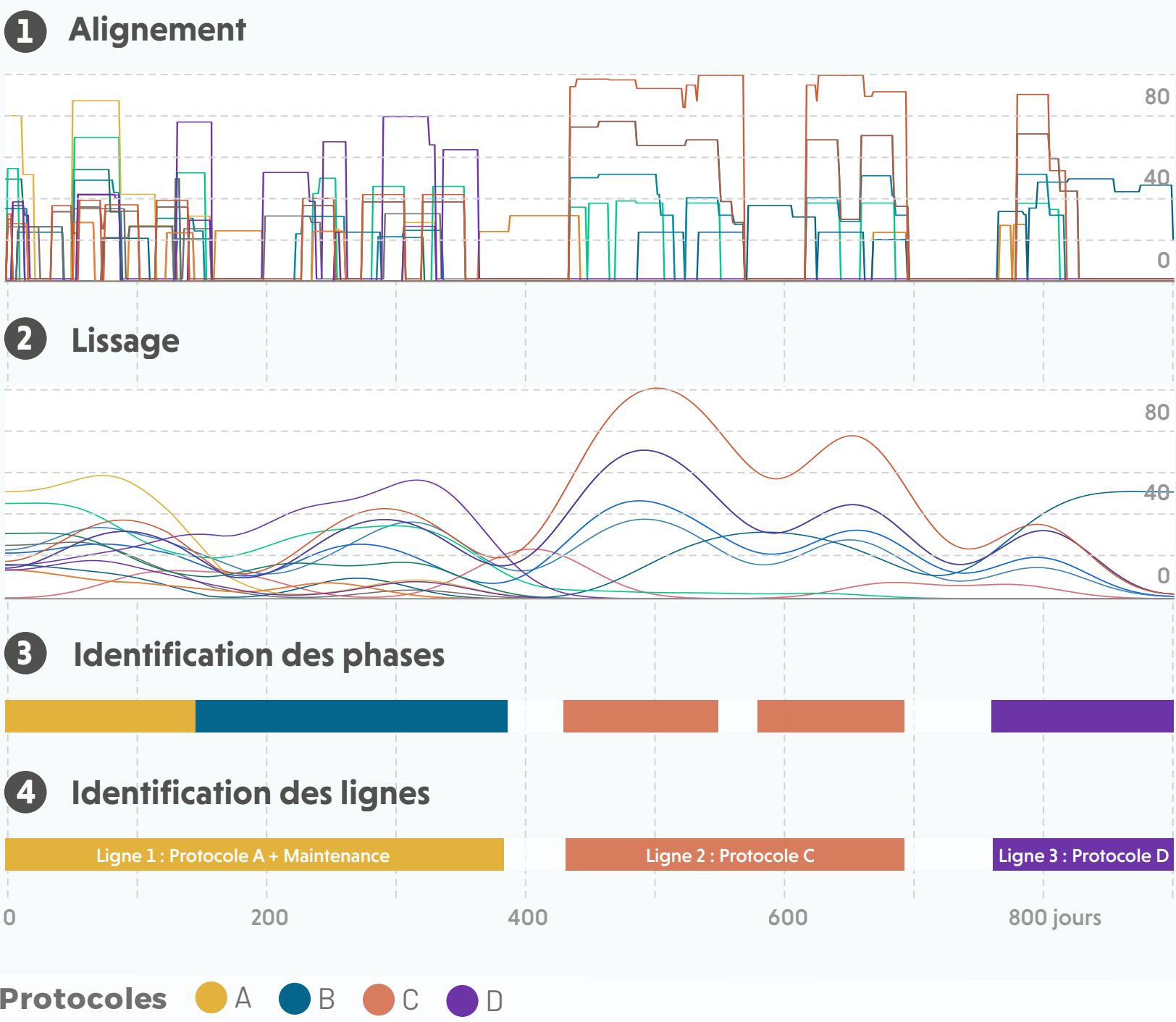
INTRODUCTION

Le **Système National des Données de Santé (SNDS)** rend accessible de manière pseudonymisée l'ensemble des données de consommations de soins des français à des fins d'analyses de santé publique. L'objectif de ce travail est de décrire les avantages et certaines limites du SNDS concernant l'identification des lignes de traitement à travers les résultats de l'étude EpiCART.

MÉTHODE

EpiCART est une étude de cohorte rétrospective incluant des patients traités pour une leucémie aiguë lymphoblastique (LAL), un lymphome B diffus à grandes cellules (LBDGC), ou un lymphome folliculaire (LF) entre 2017 (2015 pour le LF) et 2020. Seules les hospitalisations comportant un code CIM-10 de chimiothérapie, les molécules de la liste en sus, les chimiothérapies orales de ville et les fréquences d'administration sont disponibles dans le SNDS pour reconstituer les lignes de traitement de ces trois pathologies. L'algorithme d'intelligence artificielle ATLAS (Analysis of Treatment Lines using Alignment of Sequences)^{1,2} a été utilisé afin de reconstituer les lignes de traitements en comparant chaque séquence de traitement avec chaque protocole théorique renseigné, et pour calculer un score de similarité pour chacun (Figure 1). La survie globale (SG) à 1 an par ligne issue de précédents articles est présentée comme proxy de la précision de l'analyse faite par ATLAS.

Figure 1. Identification des lignes de traitement par ATLAS en 4 étapes. Pour chaque patient : (1) calculer un score de similitude avec chaque protocole théorique chaque jour du suivi et alignement de chacun de ces protocoles (2) Lisser les scores et identifier les phases (= enchaînement de cycles) sur tout le long de la séquence et (3-4) reconstituer les lignes de traitement (= plusieurs phases)



CONCLUSION

L'utilisation d'ATLAS a permis d'identifier correctement des séquences thérapeutiques au sein des patients traités pour un LF ou un LBDGC. Malgré ses performances déjà publiées, et une application cohérente de l'algorithme sur deux autres hémopathies au sein de l'étude, les résultats sur la LAL ont montré une identification des lignes de traitement insuffisante, rendant peu exploitable les résultats de SG au sein de cette cohorte. L'enrichissement du SNDS et des données intra-GHS avec le renseignement des molécules administrées permettrait d'améliorer la qualité de la génération de données de vraie vie sur plan national.

RÉSULTATS & DISCUSSION

L'utilisation d'ATLAS a permis de travailler sur l'ensemble du suivi de cohortes de taille importante (2 448 patients pour la LAL, 17 102 pour le LF, et 20 655 patients pour le LBDGC), comportant chacune un nombre conséquent de protocoles de traitement différents (33 pour la LAL, 18 pour le LF, 20 pour le LBDGC).

LYMPHOME FOLLICULAIRE

La SG à 1 an en première ligne (L1) pour le LF était cohérente avec les résultats de Rummel et al. (95 %) ainsi que Ardeshtna et al. (99 %)^{3,4}. Il en était de même avec les données de seconde ligne (L2) avec une survie à 2 ans de 93 % chez Leonard et al. et une survie à 5 ans de 94 % chez Kahl et al.^{5,6}. En revanche, les résultats retrouvés par la métaanalyse de Kanters et al. étaient inférieurs à ceux de notre étude pour la L3 avec une SG à 1 an de 75 %⁷. Cette disparité peut être expliquée par le fait que Kanters et al. ont basé leur estimation sur une L3+ (incluant les lignes subséquentes) et non une L3 stricte. Enfin, à la fois les résultats de SCHOLAR-5 et ceux de Kanters et al. étaient en accords avec ceux d'EpiCART pour la L4+ avec 72,3 % et 75 %, respectivement^{7,8}.

	Nombre de patients par an (min – max)	SURVIE GLOBALE À 1 AN (IC 95 %)	
		EpiCART	Littérature
LIGNE 1	4 631 – 4 816	96 % (96 ; 97)	95 - 99 % ^{3,4}
LIGNE 2	995 – 1 089	91 % (89 ; 92)	> 93 % (Survie Globale à 2 ans) ⁵ > 94 % (Survie Globale à 5 ans) ⁶
LIGNE 3	281 – 389	85 % (82 ; 87)	75 % ⁷
LIGNE 4+	93 – 203	72 % (66 ; 77)	72,3 % - 75 % ^{7,8}

LYMPHOME B DIFFUS À GRANDES CELLULES

Concernant le LBDGC, les résultats retrouvés étaient cohérents pour la L1 avec par exemple une SG à 1 an de 90 %⁹. Ils l'étaient moins à partir de la L2, avec des résultats retrouvés dans la littérature variant entre 65 et 74 % selon les références¹⁰⁻¹². La divergence trouvée pour la troisième ligne (45 % chez Wang et al. vs. 58 % pour EpiCART) peut en partie s'expliquer par l'ancienneté des données de la publication de Wang et al.¹³.

	Nombre de patients par an (min – max)	SURVIE GLOBALE À 1 AN (IC 95 %)	
		EpiCART	Littérature
LIGNE 1	5 257 – 5 588	88 % (87 ; 88)	90 % ⁹
LIGNE 2	1 752 – 2 043	71 % (69 ; 73)	65 - 74 % ¹⁰⁻¹²
LIGNE 3	666 – 894	58 % (55 ; 61)	45 % ¹³
LIGNE 4+	300 – 438	52 % (47 ; 56)	—

LEUCÉMIE AIGUË LYMPHOBLASTIQUE

Les taux observés pour la cohorte LAL étaient surestimés dès la L1 avec une SG à 1 an de 74 % retrouvée par Monnereau et al.¹⁴. Il en était de même avec la L2 et avec des estimations entre 24 % et 35 % (vs. 66 % dans EpiCART)¹⁵⁻¹⁸. Enfin, la surestimation était la plus importante pour la L3 avec 8 % chez O'Brien et al. (vs. 61 % dans EpiCART)¹⁹. Les écarts importants retrouvés dans l'analyse des séquences de traitement de la LAL peuvent s'expliquer par plusieurs éléments. L'identification des protocoles et des changements de lignes n'a pu s'effectuer que sur l'analyse du rythme d'occurrence des séjours hospitaliers avec un code CIM-10 de chimiothérapie et de LAL en raison de l'absence de molécule de la liste en sus (hormis le rituximab) sur notre période d'étude. Cet élément combiné à un taux de rechute élevé a eu pour conséquence de conserver à tort des patients dans une ligne N alors qu'ils avaient changé pour une ligne N+1, et ainsi surestimer la survie en ligne N. Afin de palier ce déficit de variables indicatrices, deux méthodes seraient envisageables. L'utilisation des données d'un registre, ou bien enrichir la base de données du SNDS avec les

	Nombre de patients par an (min – max)	SURVIE GLOBALE À 1 AN (IC 95 %)	
		EpiCART	Littérature
LIGNE 1	749 – 823	82 % (79 ; 84)	74 % ¹⁴
LIGNE 2	194 – 241	66 % (60 ; 71)	24 - 35 % ¹⁵⁻¹⁸
LIGNE 3	34 – 59	61 % (46 ; 73)	8 % ¹⁹
LIGNE 4+	10 – 14	NA	—

données intra-GHS comme il est actuellement expérimenté dans certains établissements français dans le cadre d'une expérimentation Article 51 en cours²⁰. La dernière option permettrait de combiner une meilleure précision concernant l'analyse des séquences de traitement à l'exhaustivité de la base de donnée.

Références :

1. Prodel M, et al. 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). 2019.

2. Laurent M, et al. Rev D'Épidémiologie Santé Publique. 2020.

3. Rummel, M. J. et al. J. Clin. Oncol. 2005.

4. Ardeshtna, K. M. et al. The Lancet Oncology. 2014.

5. Leonard, J. P. et al. J. Clin. Oncol. 2019.

6. Kahl, B. S. et al. J. Clin. Oncol. 2014.

7. Kanters, S., et al. BMC Cancer 2023.

8. Ghione P, et al. Blood. 2022.

9. Camus V, et al. Blood Adv. 2022.

10. Salles, G. et al. The Lancet Oncology. 2020.

11. Locke, F. L. et al. N Engl J Med. 2022.

12. Abramson, J. S. et al. Blood. 2022.

13. Wang, M. et al. Leukemia. 2013.

14. Monnereau A, et al. Institut national du cancer. 2021.

15. Oriol A, et al. Haematologica. 2010.

16. Topp MS, et al. Cancer. 2021.

17. Kantarjian HM, et al. N Engl J Med. 2016.

18. Kantarjian H, et al. N Engl J Med. 2017.

19. O'Brien S, et al. J Clin Oncol. 2013.

20. Arrêté du 28 août 2019 relatif à l'expérimentation faisant évoluer les modalités de la connaissance de l'utilisation et de la prise en charge des médicaments onéreux administrés par les établissements de santé - Légifrance [Internet]. Available from: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000039003584>

Source des données :

Étude SNDS enregistrée auprès du HDH le 24 mars 2021 et autorisée par la CNIL en date du 4 Mai 2021 (DR-2021-140 (demande 921191)) - Convention CNAM signée le 27 octobre 2021